GridFS in MongoDB: Study Material

1 Introduction to GridFS

GridFS is a specification in MongoDB used to store and retrieve large files such as PDFs, images, videos, and other binary data that exceed the BSON document size limit of 16 MB.

Instead of storing a file as a single document, GridFS divides the file into smaller chunks (default size: 255 KB) and stores them across multiple documents.

2 How GridFS Works

GridFS uses two collections:

2.1 1. fs.files

Stores metadata about the file:

- File name
- File size
- Upload date
- Chunk size

2.2 2. fs.chunks

Stores the actual file data in chunks:

- Each chunk is a part of the file
- Linked using `files_id`
- Ordered using chunk number `n`

3 Working Flow

1. File is split into chunks
2. Chunks stored in `fs.chunks`
3. Metadata stored in `fs.files`
4. File reconstructed during retrieval

4 Installation of MongoDB Database Tools

4.1 Linux (Ubuntu/Debian)

```
sudo apt update
sudo apt install mongodb-database-tools
```

Verify installation:

```
mongofiles --version
```

4.2 Windows

1. Download MongoDB Database Tools ZIP from official MongoDB website
2. Extract the ZIP file
3. Navigate to extracted folder:

```
mongodb-database-tools\bin
```

1. Add this path to Environment Variables (Path)
2. Restart Command Prompt

Verify installation:

```
mongofiles --version
```

5 Using GridFS with mongofiles

Upload a file:

```
mongofiles -d dhc put file.pdf
```

Download a file:

```
mongofiles -d dhc get file.pdf
```

6 When to Use GridFS

Use GridFS in the following cases:

- File size is greater than 16 MB
- Need to store large binary data
- Streaming of large files is required
- Partial file retrieval is needed

7 When Not to Use GridFS

Avoid GridFS when:

- Files are small (less than 16 MB)
- High-performance file serving is required
- Static files can be stored in file systems or cloud storage (e.g., AWS S3)

8 Advantages (Merits)

- Handles large files efficiently
- Supports file streaming
- Works with MongoDB replication and sharding
- Allows partial file access

9 Disadvantages (Demerits)

- Slower than file system storage
- More complex structure
- Increased storage overhead due to chunking

10 Best Practice

In real-world applications:

- Store files in GridFS
- Store file references (file_id) in application collections

Example:

```
db.c22.insertOne({
    name: "file.pdf",
    file_id: ObjectId("..."),
    uploaded_at: new Date()
})
```

11 Conclusion

GridFS is an effective solution for handling large files in MongoDB. It is best suited for applications requiring scalable file storage, but should be avoided for small files or high-speed static content delivery.